Behavioural Public Policy (2022), 6: 4, 554–577
© *Cambridge University Press* doi:10.1017/bpp.2019.22 First published online 18 July 2019

Difficult training improves team performance: an empirical case study of US college basketball

HEATHER BARRY KAPPES*
London School of Economics and Political Science, London, UK
ADAM L. ALTER
New York University, New York, NY, USA
GRIFFIN S. EDWARDS
University of Alabama at Birmingham, Birmingham, AL, USA
DAVID J. BERRI
Southern Utah University, Cedar City, UT, USA

Abstract: One major challenge facing policy-makers is to design education and workplace training programs that are appropriately challenging. We review previous research that suggests that difficult training is better than easy training. However, surveys we conducted of students and of expert sport coaches showed that many prescribed easy rather than difficult training for those they coached. We analyzed the performance of National Collegiate Athletic Association (NCAA) basketball teams in postseason tournaments to see whether the existing research, largely on individuals in short-term situations, would generalize to teams in the long run. Indeed, playing difficult nonconference (training) games modestly improved performance for NCAA teams in the postseason. Difficult training particularly benefitted teams that lost many nonconference games, and the effect of difficulty was positive within the range of difficulty NCAA teams actually encounter, making it clear that difficult training is superior. We suggest that our results can be generalized beyond sports, although with careful consideration of differences between NCAA basketball teams and other teams that may limit generalizability. These results suggest that policy-makers might consider amplifying the difficulty of team training exercises under certain conditions.

Submitted 7 August 2018; revised 24 June 2019; accepted 28 June 2019

* Correspondence to: Heather Barry Kappes, Department of Management, London School of Economics and Political Science, London WC2A 2AE, UK. Email: h.kappes@lse.ac.uk

https://doi.org/10.1017/bpp.2019.22 Published online by Cambridge University Press

Across many policy areas, practitioners need to design training that will best help people succeed. Governments and companies use education and training to build skills that improve performance and productivity. Maximizing the benefits of training requires identifying what kind of training builds skills most effectively. The world of work is increasingly one of teams (Deloitte, 2018), so the key policy question is what kind of training is most beneficial for team performance.

One crucial question about training is how difficult it should be. An analogy to medical vaccinations suggests that people should be exposed to milder trials (i.e., easy training) before they experience a full-blown challenge. Although early research suggested that this analogy holds for tasks that are completely novel or very challenging (Barch & Lewis, 1954), a larger body of evidence in the behavioral sciences suggests that people benefit more from particularly difficult training (Keith & Frese, 2008; Soderstrom & Bjork, 2015). However, this evidence largely stems from studies of individuals rather than teams, and looks at relatively short-term rather than longer-term outcomes.

To quantify the effects of difficult versus easy training on the real-world performance of teams over time – as opposed to performance on a short-term task or test that researchers have constructed – we looked to the domain of sports. Although sports require physical as well as intellectual skill, lessons from sports have been generalized to other settings, including employees (Katz, 2001) and students (Neri, 2017). We constructed a dataset consisting of all the nonconference (preseason), regular season and conference games over a 10-year period in National Collegiate Athletic Association (NCAA) basketball. These data allowed us to observe the effects of difficult versus easy training on performance, in a team, over time.

Difficult versus easy training

We began our research by asking what expert practitioners believed about difficult versus easy training. Speaking to an audience of approximately 300 coaches and experts at the Australian Institute of Sport in November 2016, the second author asked them, "All else being equal, do you prefer easy training sessions or harder ones?" The example of archery was used to illustrate the idea (where the training target is closer – easier – or further away – harder – than the test target), while saying that the same distinction between easy and hard training would apply to other sports like basketball, soccer, etc. To avoid herding and social demand effects, audience members were asked to close their eyes and raise a hand to show their preference for easy versus difficult training. Approximately 85% of these experts (whose responses were counted by observers as well as the presenter) said they preferred easy training. Asked for

https://doi.org/10.1017/bpp.2019.22 Published online by Cambridge University Press

clarification, coaches said they preferred the easy approach because they were concerned about exhausting or demotivating players. They only used hard training with very seasoned veterans and only infrequently.

The bulk of the evidence from the behavioral sciences does not support these coaches' preference for easy over difficult training. This evidence comes from multiple lines of research, dating back to early studies on transfer of training with Air Force personnel (Holding, 1962). Difficult tasks increase the likelihood that people will make errors, which facilitate learning because they provide informative feedback (Fisher & Lipson, 1986; Heimbeck *et al.*, 2003) and encourage people to recruit additional cognitive and physical resources to overcome those errors (Alter *et al.*, 2007). 'Error management training' incorporates this observation and finds that encouraging participants to make many errors and to learn from them produces better performance on a later focal task compared to training that prevents participants from making errors (Keith & Frese, 2008). By making errors, people improve in so-called *metacognition*, meaning that they get better at planning, monitoring and evaluating their own efforts. Although errors provide useful feedback, seeing that one is performing poorly can be upsetting. During difficult training, people gain opportunities to experience the emotions associated with making mistakes and to practice regulating them. Thereby, the enhanced emotion control that helps people to persevere in the face of challenge may account for some of the downstream benefits of difficult training (Bandura, 1997; Duckworth *et al.*, 2007). Indeed, recent research suggests that talented individuals who become top performers actually benefit from experiencing life stress, which helps with mental toughness and resilience (Collins & MacNamara, 2012); people in general report better mental health and well-being if they have experienced a moderate amount of lifetime adversity (Seery *et al.*, 2010). Finally, consider that after easy training, the demands encountered during actual performance may seem overwhelming; those same demands may seem relatively low when compared to difficult training, thereby boosting confidence and enhancing success. This is the same mechanism musicians are drawing on when they practice playing a difficult piece at double speed, or when baseball players swing two bats as a warm up. These lines of research together make a compelling case that difficult training is superior to easy training for helping people acquire new physical and intellectual skills (Bjork & Bjork, 2011; Soderstrom & Bjork, 2015). Indeed, this diverse research on training echoes the message from the goal-setting literature that shows performance benefits of very difficult goals (Garland, 1983; Locke & Latham, 1990), including in team settings (Kleingeld *et al.*, 2011).

Why, then, might expert coaches vote against difficult training? One possibility is that they recognize the benefits outlined above, but believe that these benefits

https://doi.org/10.1017/bpp.2019.22 Published online by Cambridge University Press

are canceled out by exhaustion or demotivation following difficulty. Given that people tend to interpret the way they perform during training as an indication of their competence (Soderstrom & Bjork, 2015), difficult training probably leads players to perceive themselves as less skilled, which may dampen motivation. Indeed, it may be difficult to assess the relative benefits of training regimes in the moment: difficult training, during which people perform poorly, may feel ineffective, while easy training, which produces a sense of fluency and is enjoyable, might be interpreted as effective (Sitzmann *et al.*, 2010). The net result of any trade-off between learning (enhanced) versus motivation (decreased) following difficult training cannot be quantified by previous research, because most of the research studies showing benefits of difficulty observe only short-term outcomes. If difficulty takes a physical toll over time or dampens motivation to the extent that people disengage, it may take time to see the (negative) effects on performance. That is, perhaps the coaches are right, and previous research findings should not be generalized to contexts where training is repeated over time and long-term outcomes are important. Moreover, one might ask whether difficult training becomes unhelpful in situations where success versus failure can be quantified (i.e., training games that have a winner and a loser), since failure could be discouraging and might further dampen motivation. If the benefits of difficulty are erased in situations where people fail (i.e., lose difficult training games), then, again, coaches may be right in avoiding it.

A second possible explanation for coaches' voting against difficult training is that perhaps only moderate difficulty helps performance, and more extreme difficulty would hurt performance. It may be that the research studies showing benefits of difficulty on performance have not tested conditions extreme enough to show that high difficulty can backfire. Indeed, very high levels of lifetime adversity are associated with worse mental health than the moderate levels that appear beneficial (Seery *et al.*, 2010). If extreme difficulty does hurt, then coaches might be making the right decision to avoid difficult training, given that it could be challenging to know ahead of time where the line between moderate and extreme difficulty lies.

A third possibility is that the same difficult training that benefits individual performance, as demonstrated in the research reviewed above, would hurt team performance. Resilient teams, those with "the capacity to withstand and recover from challenges, pressure, or stressors," are not necessarily composed of resilient individuals (Alliger *et al.*, 2015, p. 176), meaning that even if difficult training makes individual athletes more resilient, it might not help a team. When working as a team, confidence in one's teammates is crucial for success (Cohen & Bailey, 1997; Bloom *et al.*, 2003), and team members' beliefs that the team can collectively succeed are strong predictors of performance outcomes (Gully *et al.*, 2002). Perhaps difficult training decreases

https://doi.org/10.1017/bpp.2019.22 Published online by Cambridge University Press

confidence in teammates or in the team as a whole, thereby weakening team bonds; as noted above, confidence and efficacy may be particularly at risk in situations where teams fail (i.e., lose difficult training games). Meta-analysis finds that specific difficult goals (rather than nonspecific goals) improve group performance, but if individuals within the group have the goal of maximizing their individual performance – as may be the case for some individual elite athletes hoping to turn professional – group performance suffers (Kleingeld *et al.*, 2011). We are not aware of previous research directly comparing the effects of difficult versus easy training (which is different from difficult versus easy goals) on team performance, but if teams respond differently to difficult training than individuals, then coaches who lead teams would be correct in avoiding difficult training. This is an important question for policy-makers who make decisions about the workplace because most employees today work as part of a team rather than strictly individually (Deloitte, 2018).

Whereas the three lines of thought outlined above are reasons that expert coaches might be correct in preferring easy to difficult training, it is also possible that the expert coaches we surveyed are simply mistaken. In this case, it would be important to know whether this is an error they make only when designing training for others or whether they would be similarly incorrect if designing training for themselves. People sometimes prescribe less optimal actions to others than they would to themselves (Hsee & Weber, 1997; Zikmund-Fisher *et al.*, 2006), and coaches might be overly influenced by players' enjoyment of easy training. Perhaps coaches are actually recommending less effective (easier) training to their athletes than they would prescribe for themselves. To further explore the beliefs about the relative effectiveness of easy and difficult training and to investigate whether recommendations are different when made in the role of coach, we conducted a survey experiment. We asked a sample of student respondents that comprised both inexperienced athletes and elite competitors for their recommendations about training: (1) to meet goals of motivation versus preparation; (2) in an individual versus team sport; and (3) for themselves versus for others (i.e., in the role of coach). We varied these three factors – the former two within respondents and the latter between respondents – to see how these factors affected the recommended training.

Survey experiment

Methods

Participants and design

Participants were 173 members of an undergraduate subject pool at a large American university who answered these questions as part of a longer

https://doi.org/10.1017/bpp.2019.22 Published online by Cambridge University Press

experimental session. Gender and age were not recorded, but participants were asked about their experience in both individual and team sports; 74% reported that they had competed in sports, including 32% at an elite level.

The study used a 2 (goal: motivation versus preparation) × 2 (sport: running/individual sport versus basketball/team sport) × 2 (recommendations: for self versus as coach) mixed design. The first two factors were manipulated within subjects and the third factor was manipulated between subjects. The data for this and the next study are available at https://osf.io/cr8wx/.

Procedure and measures

All participants answered four questions, two about running (*individual sport condition*) and two about basketball (*team sport condition*). For each sport they were asked separately about which strategy would be most useful for motivating further training (*motivation goal condition*) and which strategy would be most useful for training to do well (*preparation goal condition*). Half of the participants ($n = 86$) were randomly assigned to make recommendations for themselves; the other half ($n = 87$) to make recommendations in the role of a 'coach'. We manipulated the target of the recommendations between rather than within subjects to prevent participants from anchoring on one set of recommendations when making the other – that is, to give the best opportunity to observe whether people were inclined to recommend different training for themselves versus for those they were coaching.

Participants were instructed, regarding running: "Let's say you'd like to compete in a race in a few months' time…" (*recommendations for self condition*) or "Please imagine that you are a coach being asked to advise some athletes about how to prepare for upcoming competition. Let's say you're coaching a runner who would like to compete in a race in a few months' time…" (*recommendations as coach condition*). Participants were then asked: "Which of the following strategies will be most useful for motivating yourself [the runner] to train for the race?" and "Which of the following strategies will be most useful for training yourself [the runner] to run quickly in the future race?" For both questions, participants chose between "Going out and running an extremely difficult training run," "Going out and running a very easy training run," and "These are equally useful." Questions about basketball followed a parallel format.

Results

We recoded responses so that choosing difficult training was coded as 1 and choosing easy training or saying the two strategies were equally useful was

https://doi.org/10.1017/bpp.2019.22 Published online by Cambridge University Press

coded as 0,[1] and analyzed the responses using a 2 (goal: motivation versus preparation) × 2 (sport: running/individual versus basketball/team) × 2 (recommendations: for self versus as coach) mixed design. We used binary logistic regression analysis with generalized estimating equations to account for correlations between the multiple responses submitted by each individual. In our first model, the probability of choosing difficult training versus another outcome was predicted by the main effect of each factor, all two-way interactions and the three-way interaction (goal × sport × recommendation target).

Three effects were significant at the $p < 0.10$ level in this model. First, there was a large effect of goal (Wald $\chi^2(1) = 58.97$, $p < 0.001$). As is seen in Figure 1, participants were much more likely to recommend difficult training when the goal was to choose the training that would be best preparation for future performance as opposed to motivating people to train further. Second, there was an effect of recommendation target (Wald $\chi^2(1) = 2.90$, $p = 0.089$). The probability of recommending difficult training was higher when making recommendations for self (0.56) than when making recommendations as a coach (0.47). Finally, there was an interaction effect of sport × goal (Wald $\chi^2(1) = 3.28$, $p = 0.07$). As is seen in Figure 1, participants were slightly more likely to choose difficult training as the more motivating and slightly less likely to choose it as the more helpful when making recommendations for the team sport (basketball) rather than the individual sport (running). The three-way interaction (goal × sport × recommendation target) was not significant (Wald $\chi^2(1) = 0.001$, $p = 0.98$), and none of the other main effects or interaction effects in the model approached significance (p-values > 0.38).

A further analysis showed that, although respondents who had experience competing in sports at an elite level (approximately 32% of the respondents) were more likely to recommend difficult training than respondents without such experience (0.58 versus 0.48; Wald $\chi^2(1) = 3.24$, $p = 0.072$), there was no evidence that elite experience affected how participants responded to the other manipulations. That is, none of the two-way interaction effects (elite

1 We used this approach because we were specifically interested in factors that affect the likelihood of perceiving difficult training as the most useful. Across questions, between 10% and 21% of respondents indicated that the difficult and easy training were equally useful. If we treat "they are equally useful" responses as missing and compare only recommendations for difficult versus easy training, conclusions are identical regarding the effect of goal ($\chi^2(1) = 36.34$, $p < 0.001$) and the effect of recommendation target ($\chi^2(1) = 2.97$, $p = 0.085$). However, the interaction effect of sport × goal is no longer statistically detectable ($\chi^2(1) = 0.20$, $p = 0.66$). Thus, readers should bear in mind that our inferences about differences in the recommendations that survey respondents made about running (an individual sport) versus basketball (a team sport) rely on comparing the belief that difficult training is superior to all other beliefs.

https://doi.org/10.1017/bpp.2019.22 Published online by Cambridge University Press

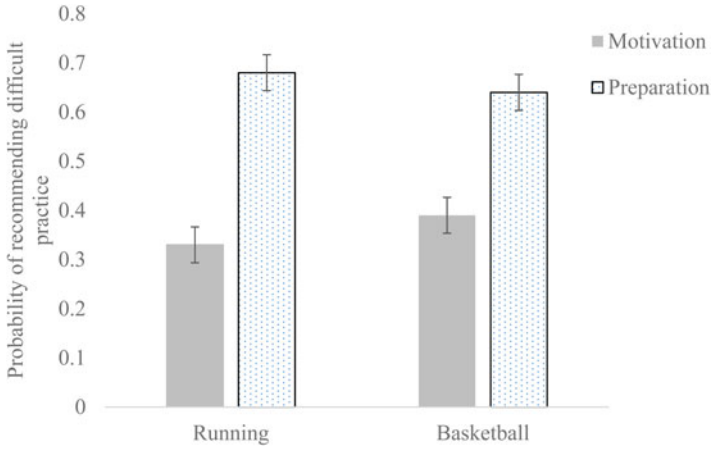


Figure 1. The probability of recommending difficult practice in the survey experiment as a function of the goal of the training (motivating people to continue versus preparing people to succeed) and the type of sport (running/individual versus basketball/team). Error bars are ± 1 standard error.

experience by goal, by sport or by recommendation target) were statistically significant (p-values > 0.24).

Discussion

The results of the survey experiment speak to our introductory speculations as to why coaches might hesitate to recommend difficult training. First, participants strikingly preferred easy training for motivating athletes, but difficult training for preparing them to perform well in the future; we had conjectured that this trade-off might be one reason as to why expert coaches chose easy over difficult training. It seems that people more generally, not only experts, suspect that there is a trade-off between these two effects of difficult versus easy training. Second, we saw that, perhaps surprisingly, people were more likely to prescribe difficult training for themselves than in the role of a coach. Although speculative, given the small and non-representative sample in our survey experiment, the difference between recommendations for self versus for others could explain why most expert coaches we surveyed earlier did not recommend difficult training to the athletes they coached. Lacking certainty about the athletes' (in the survey question) level of commitment, perhaps coaches were too hesitant to use the method that they probably recognize (as our survey respondents did) produces superior performance, even if they would have been willing to use this method themselves.

https://doi.org/10.1017/bpp.2019.22 Published online by Cambridge University Press

Finally, we observed some differences in the recommendations that survey respondents made about running (an individual sport) versus basketball (a team sport). While acknowledging that there are other characteristics of these two sports that could also account for the differences, one interpretation is that some respondents believe difficult training is slightly less beneficial as preparation (while still being less motivating than easy training) in a team than an individual sport. If these beliefs are correct, then the behavioral sciences research that finds difficult training to be more effective may not apply to teams. We turned to our NCAA basketball data to see if this is true of real performance over time.

NCAA basketball games data

To quantify the effects of difficult versus easy training on the real-world performance of teams, we constructed a dataset consisting of all the nonconference (preseason, which we treat as training), regular season and conference games over a 10-year period in NCAA basketball. This dataset allowed us to observe the effects of difficult versus easy training on performance in a team over time.

In addition to the simple linear effect of training difficulty, we studied two further questions. First, we followed up the linear effect by testing for nonlinearity in order to see whether the degree of difficulty – from moderate to extreme – matters. Second, we examined whether teams responded differently depending on whether they won or lost their nonconference (training) games. Given the secondary nature of the data, we are not able to directly observe the preceding effects, but our analyses do allow us to infer whether difficult training produces superior performance because it: (1) simply sharpens skills (helping regardless of outcome); (2) builds confidence only when teams win in the face of considerable difficulty (helping only when teams win); or (3) builds coping skills in the face of failure (helping only when teams lose). The results of these two sets of further analyses testing nonlinear effects and testing moderation by training game outcomes allow us to refine our recommendations about easy versus difficult training for teams.

Methods

Data

We used a dataset with NCAA college basketball teams from the big five conferences for the ten seasons from 2003–2004 to 2012–2013 ($n = 605$). Those

https://doi.org/10.1017/bpp.2019.22 Published online by Cambridge University Press

conferences include the Atlantic Coast Conference, Big 12, Big East, Pac-12 and Southeastern Conference, and team performance is measured at the team by year level.

College basketball has nonconference games (like a preseason, which we treat as training games), a regular season (conference games) and a postseason. In the preseason, teams play nonconference games against teams that are not in their regional grouping, so results do not affect their conference ranking. In the postseason, each team plays a knockout tournament against the other teams in its conference, and the strongest teams then go on to compete in the NCAA tournament. We mapped the questions outlined above onto this dataset by asking: To maximize tournament success, would it be better for a team to play nonconference games that are relatively difficult or relatively easy? Specifically, we tested for an effect of nonconference play on postseason performance to ascertain whether a relatively strong (or weak) nonconference schedule had any statistical impact on postseason outcomes, and if so, whether that impact was positive or negative.[2]

To capture the difficulty of each team's nonconference (i.e., training) schedule, we rely on a measure known as Simple Rating System (SRS). SRS calculates a team's quality by considering not only that team's performance, but also the performance of the teams it plays (Drinen, 2006). More specifically, it considers a team's average margin of victory, as well as the average margin of victory of the team's opponents. Although the calculation of SRS requires one to solve a system of equations, it is approximated by knowing a team's point differential and strength of schedule (SOS), or the combined winning percentages of the opponents the team played. For example, the 2012–2013 Florida team that made it to the Elite Eight in the NCAA tournament had a point differential (average margin of victory during the season) of 17.0. This team's SOS, or the combined winning percentages of the opponents Florida played, was 6.8, so Florida's SRS score was 23.9 (there is some rounding).[3] That same year, Auburn, which lost the first game of the conference tournament and did not play in the NCAA tournament, had a point differential of –4.0 and an SOS of 4.2, so Auburn's SRS score was 0.2. As Drinen (2006) notes, the interpretation of SRS is relatively easy: "If Team A's rating is 3 bigger than Team B's, this means that the system thinks Team A is 3 points better than Team B." Most schools in major conferences typically have an SRS greater than zero

2 Another outcome one might consider is regular season performance. Regular season games are not a good outcome because they vary from conference to conference. Even among the top conferences, some conferences in any given year are harder than others. The NCAA tournament is the same for everyone, so that gives us the same test for teams even in different conferences.

3 See https://www.sports-reference.com/cbb/conferences/sec/2013.html.

https://doi.org/10.1017/bpp.2019.22 Published online by Cambridge University Press

because their SOS values are relatively high. It is not uncommon, though, for a mid-major school to have a negative SRS score.

To calculate the quality of the nonconference opponents, we take the average SRS score of the nonconference teams played. Average nonconference SRS scores in our sample ranged from –7.0 to + 11.6 (M = 0.81, SD = 3.13), where higher scores indicate a more difficult preseason schedule. Nonconference schedules are arranged by each school's athletic director and coaches. While the games are not randomly assigned, they are often decided several years in advance. Since nonconference games are scheduled years in advance, long before the composition and strength of each team is known, the difficulty of nonconference games is largely exogenous.[4] Additionally, unlike the National Football League (NFL), where players cannot compete until three years after they have graduated from high school, the National Basketball Association (NBA) allows players to compete just one year after high school graduation. The turnover, particularly among star high school players, is therefore considerable, making it especially hard to predict the strength of any particular team even one year in advance, thus making it difficult to try and tailor the nonconference schedule to the quality of the team.

To measure postseason success, we wanted to ensure all of the relevant variation was exploited. To do this, we measured tournament success by counting the number of games played and adding one to the team that won the tournament. We do this separately for the knockout tournament that each team plays against the other teams in their conference (i.e., conference tournament) and for the NCAA tournament. Historically, the NCAA tournament consisted of 64 teams, giving the tournament six rounds of games. In 2001, though, a 65th team was added, which gave the tournament a play-in game between two teams prior to the traditional six rounds. In 2011, three more teams were added, and the number of play-in games was increased to four. The teams selected for this tournament consist of the winners of 32 Division 1-A conferences. The remaining teams needed for the field are then selected by a committee. Ideally, these 'at-large' teams are the 'best' teams that failed to win an automatic conference bid.[5] Our measure of games played in the tournament plus one for the winner allows us to identify all of the teams that qualified for the tournament (which, in the case of the NCAA tournament for some

4 As an example, Brigham Young University basketball had, prior to the 2017–2018 season, 100% of the 2018–2019 nonconference games scheduled, 90% of the 2019–2020 games scheduled and 81% of the 2020–2021 nonconference season scheduled.

5 See http://www.ncaa.com/news/basketball-men/article/2017-03-12/march-madness-bracket-how-68-teams-are-selected-division-i.

https://doi.org/10.1017/bpp.2019.22 Published online by Cambridge University Press

schools, may, in it of itself, be considered a success) and still acknowledge the tournament winner.[6]

To control for potential confounds, we extracted a range of variables from each team's performance during its in-season games. Those variables were each team's own regular season win/loss record and average point margin (see Pitts, 2016, for a similar approach) and team (or conference) fixed effects and season fixed effects. The team fixed effect variable is a dummy variable for each team. Each team's fixed effect captures factors that influence performance and remain relatively stable across time, including brand power, loyalty of fan base and quality of training facilities and recruiting and coaching staff. This set of dummy variables captures any idiosyncratic differences between the teams – including those mentioned above – that tend not to change over time. Similarly, certain model specifications include conference fixed effects in place of team fixed effects. Whereas team fixed effects absorb the time-invariant idiosyncratic differences between teams, conference fixed effects account for the time-invariant differences between conferences. Conference fixed effects would be appropriate if there were reason to believe that there was some sort of unobserved factor correlated with conferences that was driving the results – that is, if it were not so much the team unobservables that drive wins, but rather the unobserved quality/heritage/reputation of the conference. We report our main results with models that include both of these controls to show the insensitivity of the results to the level of fixed effects. Additionally, a season fixed effects variable – a dummy variable for each season – captures unobserved factors that are common to all teams and vary across time, including the tendency for players to improve as they become more experienced.

Model validity

There are a couple of challenges regarding identification when trying to model the run a NCAA team makes in a tournament. Of primary concern is the nature of the dependent variable, which creates some challenges for ordinary least squares (OLS) regressions. The data are discrete in nature – a team cannot win, for example, 1.23 games in either tournament. Additionally, the range of possible outcomes extends from zero to six or seven. Given the

6 While alternative measures of tournament success exist, each has its shortcomings. A simple count of games played in the tournament fails to recognize the tournament winner, and a straight count of games won in the tournament would treat the teams that earned a spot in the tournament but lost the first game identically to the teams that failed to reach the tournament. Doing the latter would leave almost half of the variation in our data misspecified. While we choose the number of games played plus one for the tournament winner as our preferred outcomes, the results are completely robust to the alternative specification where we count only games played.

https://doi.org/10.1017/bpp.2019.22 Published online by Cambridge University Press

cutoffs at zero and seven and the discreteness of the data, OLS is probably an inappropriate modeling technique. Since the data are discrete in nature and truncated at both ends, they show a distribution that is more similar to a Poisson than a normal distribution, and so we estimate every specification using Poisson regression techniques with quasi-maximum likelihood (QML) estimation.[7] Gourieroux *et al.* (1984) prove that the QML estimator is consistent and asymptotically normal when the Poisson model is correctly assumed, and Wooldridge (1999) shows that only correct specification of the condition mean is necessary for consistency. This allows the model to be distributionally misspecified, without regard to over- or under-dispersion, and still to get consistent and efficient results. This method is especially useful in this study, as it allows for robust standard errors clustered at the state level (Wooldridge, 1999).

Another potential challenge to identification is the host of unobserved variables that may influence both our outcomes of interest and the difficulty of each team's nonconference schedule (average opponent SRS). One advantage of the data in this domain, however, is that teams have very limited control over the difficulty of their nonconference schedules. As an example, there is very little correlation between a team's previous year's tournament performance and the next season's nonconference SOS. The unadjusted correlation between the previous year's performance and next season's average nonconference opponent SRS is 0.24 for the conference tournament and 0.34 for the NCAA tournament. Still, certain teams have a heritage of success in basketball, so to control for the possibility that stronger teams were more likely to construct difficult preseasons, each model includes team fixed effects to capture any non-observed, time-invariant factor that makes some teams more successful than others. The model proposed by Hausman *et al.* (1984) suits this issue well as it allows for fixed effects.[8]

Model

The baseline model we estimate is a Poisson fixed-effects model that imposes a linearity assumption between SRS and tournament run. Generally, we estimate:

$$r_{it} = m(\alpha + \beta s_{it} + X_{it} + F_i + Y_t + e_{it}) \quad (1)$$

7 We replicate our main results using OLS and find no major difference relative to the Poisson regressions in the main results.

8 Since our variable of interest – average opponent SRS – varies within group, we chose a fixed-effects Poisson model over a random-effects model, as the fixed-effects model does not carry the assumption that the unobserved error is not correlated with any of the independent variables. The results are insensitive to random-effects specifications.

https://doi.org/10.1017/bpp.2019.22 Published online by Cambridge University Press

where r is the count of the tournament run either in the conference tournament or the NCAA tournament that varies by team, I, in season, t, α is the intercept, X is a vector of covariates that include regular season win/loss ratio and average point margin that vary by team and season, F is a matrix of either team or conference fixed effects (depending on the specification), Y is the matrix of season fixed effect and e is the error term. The independent variable of interest s_{it} is the measure of the average SOS of the nonconference opposition for each team in each year as measured by SRS. In this approach, we allow $m(\cdot)$ to take the form $exp(\cdot)$ as proposed by Hausman and colleagues (1984), and we use QML estimation to estimate a fixed-effects Poisson regression.

In addition to the specification in equation (1), we also relaxed certain assumptions to more directly address two behavioral questions. The first revised equation allows us to examine whether the benefits of difficult nonconference schedules diminish or reverse when the nonconference schedule is exceedingly difficult. To address this question, we revise equation (1) by including a quadric term in SRS.[9]

$$r_{it} = m(\alpha + \beta_1 s_{it} + \beta_2 s_{it}^2 + X_{it} + F_i + Y_t + e_{it}) \tag{2}$$

We thus measure the degree to which there are any nonlinearities by performing a test of joint significance on the linear and quadratic terms of s_{it}.

The second revised equation allows us to examine whether teams benefit from difficult nonconference schedules differently depending on whether they win or lose those preseason games. While beating a strong nonconference opponent may boost confidence, teams might also acquire technical and psychological coping skills when they lose difficult nonconference games. To examine this question, we interact the SRS variable with the demeaned preseason win percentage, or:

$$r_{it} = m(\alpha + a_1 s_{it} + a_2(nwr_{it} - \overline{nwr}) + a_3 s_{it} \times (nwr_{it} - \overline{nwr}) + X_{it} + F_i + Y_t + e_{it}) \tag{3}$$

where *nwr* is the nonconference win rate. In this specification, the marginal effect of s_{it} at the mean nonconference win rate is a_1. The coefficient a_3 measures the effect of playing tougher preseason games *and* winning those games – that is, if a_3 is indistinguishable from 0, then winning versus losing nonconference games does not change the effect of preseason difficulty on

9 Our models suggest a quadratic relationship to be most appropriate. This is confirmed by semiparametric regressions where the team and season fixed effects are parametrically modeled and the relationship between SOS and tournament run is allowed a flexible nonparametric form. The semiparametric models confirm the appropriateness of a quadratic relationship.

https://doi.org/10.1017/bpp.2019.22 Published online by Cambridge University Press

postseason success, but if a_3 has a nonzero value, then its direction is informative about whether winning versus losing magnifies the effect of nonconference difficulty. A positive value on a_3 suggests that the more successful teams are against nonconference opponents, the stronger the effect of preseason difficulty (i.e., SRS has a bigger effect when more games than average are won) and a negative value on a_3 suggests the opposite: that the less successful teams are in nonconference play, the stronger the effect of nonconference difficulty (i.e., SRS has a bigger effect when more games than average are lost). Alternatively stated, a_3 measures the returns to scheduling harder nonconference games for different types of teams. If a_3 is negative, then the value to scheduling tougher nonconference games is greater for teams that do not start out the season as well compared to teams that win most of their nonconference games.

Equation (3) represents a slight departure from our previous models, as we aim to measure directly the effect of winning versus losing nonconference games. Whereas previously in Tables 1–3 we include as a control the team's win ratio in order to address how winning/losing nonconference games affects postseason performance, we include only the win ratio for nonconference games in Table 4 as part of the necessary controls to measure this interactive effect.

Results

The effect of a harder nonconference schedule is reported in Table 1 for the conference tournament and in Table 2 for the NCAA tournament. Column (1) in each table reports the results with no controls or fixed effects. Column (2) includes controls and team and season fixed effects. Column (3) includes controls and fixed effects for conference and season. Every coefficient is reported as a marginal change and robust standard errors are in parentheses.

Across all specifications, both Tables 1 and 2 show a positive effect of nonconference opponent SRS on tournament run. It is important to note that this result is robust to the type of tournament. Though they are very different tournaments – every team gets into the conference tournament while only a select few in each conference make the NCAA tournament – we see that tougher nonconference competition yields a similar effect in both the NCAA and conference tournaments. That is, as teams play harder training games, they tend to make it farther in each respective postseason tournament. The coefficients reported in each regression are marginal effects that suggest how many additional games a team would play in each tournament if they increased their nonconference opponent SRS score by 1. For instance, in column (1) of Table 2, if a team increased its average opponent SRS score by 3, we would anticipate that team would advance an additional $3 \times 0.147 = 0.441$ games in the NCAA tournament. Keep in mind that these regressions control for team effects and the

https://doi.org/10.1017/bpp.2019.22 Published online by Cambridge University Press

Table 1. Nonconference opponent quality and conference tournament run.

	(1)	(2)	(3)
Average opponent SRS	0.057***	0.024	0.020*
	(0.013)	(0.015)	(0.011)
Fixed effects			
Team		×	
Conference			×
Season		×	×
Observations	605	605	605
(Pseudo-)R^2	0.01	0.06	0.05

Notes: Each observation is measured at the team/year level. The reported coefficients are marginal effects and robust standard errors are reported in parentheses. Each regression includes controls for the win/loss ratio and point margin. The outcome is measured in number of tournament games played plus one extra for the tournament winner.
*p < 0.10, **p < 0.05, ***p < 0.01.
SRS = Simple Rating System.

win/loss ratio for each team. Accordingly, the effect we estimate suggests what would happen for the same team with the same regular-season record, which may, in part, explain why the effect is not larger.

We provide some specific examples to give a feeling for the magnitude of these effects. In the 2009–2010 season, all else being equal, Oregon (average preseason SRS = –0.53, 33rd percentile) would have advanced half a game further in the NCAA tournament had they played a more difficult preseason schedule similar to Arizona's (average preseason SRS = 7.09, 97th percentile). In the 2012–2013 season, Georgetown entered the NCAA tournament as number 2 seed and ranked number 5 in the Associated Press Poll. They were upset in the first round of the tournament, losing to Florida Gulf Coast by 10 points. Georgetown's preseason schedule consisted of mostly mid-major teams with weak SRS scores (average opponent SRS of –0.198, 38th percentile). If they had replaced half of their weaker nonconference games against mid-major teams and instead played Duke, Virginia Commonwealth University, Baylor, Brigham Young University, Alabama and Wisconsin, their average opponent SRS would have increased to 12.2, and we estimate their tournament run would have extended to the Sweet Sixteen instead of losing in the first round.

In contrast, the 2006–2007 Tennessee Volunteers made it to the Sweet Sixteen before losing by 1 point to eventual runners-up Ohio State. That year, Tennessee played the second most challenging nonconference schedule in their conference with an average opponent SRS score of 3.96 (84th percentile and second to only

https://doi.org/10.1017/bpp.2019.22 Published online by Cambridge University Press

Table 2. Nonconference opponent quality and National Collegiate Athletic Association tournament run.

	(1)	(2)	(3)
Average opponent SRS	0.147***	0.077***	0.080***
	(0.020)	(0.020)	(0.016)
Fixed effects			
Team		×	
Conference			×
Season		×	×
Observations	605	605	605
(Pseudo-)R^2	0.06	0.42	0.38

Notes: Each observation is measured at the team/year level. The reported coefficients are marginal effects and robust standard errors are reported in parentheses. Each regression includes controls for the win/loss ratio and point margin. The outcome is measured in number of tournament games played plus one extra for the tournament winner.
*p < 0.10, **p < 0.05, ***p < 0.01.
SRS = Simple Rating System.

Table 3. Nonlinear relationship between average opponent Simple Rating System score and tournament run.

	C tournament +	NCAA +
	(1)	(2)
Average opponent SRS	0.032*	0.086***
	(0.017)	(0.030)
Average opponent SRS^2	–0.004	–0.002
	(0.003)	(0.004)
P-value of joint significance	0.14	0.00
Observations	605	605
(Pseudo-)R^2	0.06	0.42

Notes: Each observation is measured at the team/year level. The reported coefficients are marginal effects and robust standard errors are reported in parentheses. The outcome is measured in number of tournament games played plus one extra for the tournament winner. Each regression includes team and season fixed effects, a control for the win/loss ratio and point margins.
*p < 0.10, **p < 0.05, ***p < 0.01.
NCAA = National Collegiate Athletic Association; SRS = Simple Rating System.

Kentucky in the Southeastern Conference, who played nonconference games averaging an SRS score of 8.29). If Tennessee had, instead, played the nonconference schedule of cross-state rival Vanderbilt (average opponent SRS = –0.84,

https://doi.org/10.1017/bpp.2019.22 Published online by Cambridge University Press

Table 4. Interactive effects between Simple Rating System and nonconference win rates.

	C tournament + (1)	NCAA + (2)
Average opponent SRS	0.025*	0.107***
	(0.015)	(0.024)
Demeaned NWR	–0.453	3.617***
	(0.505)	(0.697)
Average opponent SRS × NWR	–0.081	–0.216*
	0.075	0.113
Observations	605	605
(Pseudo-)R^2	0.06	0.43

Notes: Each observation is measured at the team/year level. The reported coefficients are marginal effects and robust standard errors are reported in parentheses. The outcome is measured in number of tournament games played plus one extra for the tournament winner. Each regression includes team and season fixed effects, a control for the win/loss ratio and point margins.
*p < 0.10, **p < 0.05, ***p < 0.01.
NCAA = National Collegiate Athletic Association; SRS = Simple Rating System; NWR = nonconference win rate.

28th percentile), we estimate their tournament would have ended a game earlier, losing to higher-ranked Virginia.

In Table 3, we examine the nature of the nonlinear relationship between nonconference opponent quality and tournament run. Each column in Table 3 represents a distinct regression with a different outcome variable. Every regression includes controls and team and season fixed effects and reports the p-value of a joint test of significance on the level and squared average opponent SRS score. As shown in Table 3, the nature of the nonlinear relationship seems to be consistent across tournaments – a positive relationship increasing at a decreasing rate – though the relationship is statistically weaker in the conference tournament with a joint test p-value of 0.14, and while the squared term in the NCAA model has a large standard error, a test of joint significance between the level and squared term suggests that, taken as a pair, they jointly matter to the model. That is, playing tougher teams is better, but at a decreasing rate. Said another way, the effect of playing increasingly tough nonconference competition levels off as the teams become increasingly strong.[10] Based on the

10 While we present these results with parametrized nonlinearities, semiparametric regressions that allow for full functional form flexibility between tournament run and preseason opponent SOS report the same quadratic result.

https://doi.org/10.1017/bpp.2019.22 Published online by Cambridge University Press

results in Table 3, we estimate that the optimal level of nonconference difficulty is a set of teams with an average SRS score of 18.8, beyond which point increasing difficulty begins to hurt performance. It is worth noting that no team in our dataset plays a nonconference schedule this difficult (the maximum is 11.6 played by Kansas in the 2004–2005 season). Theoretically, teams *could* schedule tougher nonconference games so that their average opponent strength approaches an SRS score of 18 since, typically, the top four or five teams in each major conference average an SRS score at or around 18, but in practice this is highly unlikely.[11] Thus, we can say that within the range of observed and likely nonconference schedules, increasing the difficulty of nonconference schedules helps performance.

In Table 4, we address the degree to which the outcomes of those nonconference games matters. Table 4 is organized similarly to Table 3. The coefficient of particular interest in Table 4 is the interaction between nonconference opponent strength and the nonconference win rate. Across all specifications, the coefficient is negative, suggesting that playing tougher teams and losing actually provides more benefit for both tournament runs than playing tougher teams and winning. Consistent with work such as error management theory and acquiring the skills of emotion control, this result indicates that the lessons learned from losing rather than winning tough early games to strong teams pay off in the long run. The combined results of the analyses presented in Tables 3 and 4 are approximately visualized in Figure 2.

General discussion

In brief, what we found in this archival data allows us to suggest that skilled performers working in teams should use difficult training. The benefits that this training provides enhance performance over time. Whereas coaches – and those such as managers who make decisions about training teams of people other than themselves – might be inclined to err on the side of caution by prescribing easy training, doing so predicts worse team performance in the long term.

Earlier in the paper, we identified four possible reasons why expert coaches might prefer easy training. The first three were reasons why these coaches might be correct in their preference for easy training; our results argue against these three explanations. First, we speculated that

11 For example, the top five teams in the 2006–2007 Atlantic Coast Conference averaged an SRS score of 19.1.

https://doi.org/10.1017/bpp.2019.22 Published online by Cambridge University Press

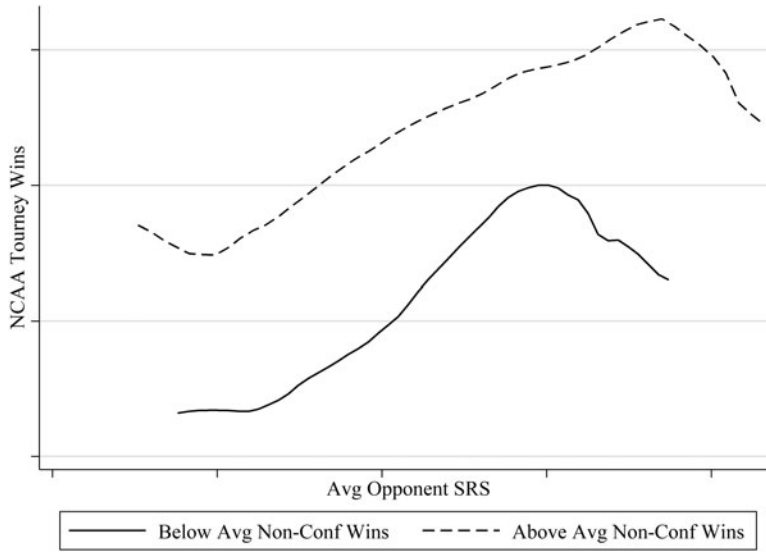


Figure 2. The nonlinear effects of difficulty of nonconference games (SRS) and the extent to which these effects vary depending on the results of the nonconference games in the archival data.

Notes: Because this figure visualizes the results of two separate sets of analyses (see Tables 3 and 4), this figure is only an approximation. For precise estimates of the magnitude of the relevant relations, see the relevant tables. The scale on the x-axis is unlabeled here to prevent misinterpretation.

NCAA = National Collegiate Athletic Association; SRS = Simple Rating System.

over time – over a longer period than has been examined in previous research – the benefits of difficulty might be counteracted by exhaustion or demotivation. Observing performance over an entire season, we find no support for that explanation; any exhaustion or demotivation is more than compensated for, since there is a positive effect of difficult nonconference games on postseason performance. Also relevant to this explanation, we examined whether teams responded differently depending on whether they won or lost their nonconference (training) games. In fact, we found that difficult nonconference (preseason) games predicted better tournament performance in general, and this effect was most pronounced among teams that lost many of their nonconference games. Thus, a fear that failure (i.e., losing games) – perhaps by discouraging players or working against team cohesion – would, over the long term, erase the benefits of difficult training is also unfounded. Competitors are perhaps more resilient than coaches assume.

https://doi.org/10.1017/bpp.2019.22 Published online by Cambridge University Press

Second, we considered the possibility that only moderate difficulty as opposed to great difficulty helps performance. To address this explanation, we tested a nonlinear effect of difficulty to see whether difficulty above a certain level begins to hurt performance. This was not the case: the benefits of difficulty level off but do not backfire until extremely high levels of difficulty – levels above what even the most challenged teams in our dataset play. Thus, the desire to avoid extreme difficulty would not be a valid reason for coaches to prefer easy to difficult training.

Third, we considered that perhaps since previous research mainly examines effects of difficult training on individuals, difficulty might not have the same benefit for teams. Although we cannot say whether effects on teams are smaller than on individuals, our data show a clear benefit to difficult training for teams, speaking against this explanation. Indeed, one might have wondered whether losing difficult training games hurts later team performance by making it harder for players to trust their teammates – in line with this speculation, we observed in our survey experiment a slight drop in recommendations for difficult training for a team sport compared to an individual one. However, as noted above, losing difficult training games actually amplified the benefits of difficult training.

On the whole, our results speak in favor of difficult training for teams when optimal performance over time is the objective. These results add to the body of research in behavioral sciences, mostly with individuals and short-term outcomes, showing the benefits of difficult training. However, it is worth considering some limitations to the generalizability of our results. One strength of this research is that preseason game difficulty and postseason performance are objective indicators. However, difficult training games are not necessarily synonymous with difficult training in general (e.g., many hours of heavy weightlifting), and we cannot speak to the relative effects of other sorts of difficulty in training. It is feasible that difficult preseason games help in part by giving players a reason to try, which might otherwise be missing in this low-stakes setting; while consistent with the arguments we have outlined about why difficulty helps, this mechanism might mean that other forms of difficult practice do not show the same effects. Moreover, NCAA seasons have limited time frames; it is possible that difficult training would be especially beneficial for teams that perform in short or seasonal bursts, whereas employees might be at heightened risk of burnout when faced with protracted extreme challenges (Sitkin *et al.*, 2011). It is also possible that teams without expert coaches, where the time for feedback is limited, or where players know that coaches themselves set the difficulty of the training schedule would not show the same benefits of difficulty. Without further research to test the effects of difficulty in teams under these different variations, one must still

https://doi.org/10.1017/bpp.2019.22 Published online by Cambridge University Press

exercise caution in designing training, even while recognizing the benefits of difficult training regimens.

Conclusion

Our investigation was motivated by the observation that although research in the behavioral sciences shows that difficult training is superior, this evidence does not appear to have been fully incorporated into practice. In the domain of sports, most of the expert coaches we surveyed preferred easy to difficult training. Although sports require physical as well as intellectual skill, lessons from sports have been generalized to groups including employees (Katz, 2001) and students (Neri, 2017), where optimal training allows for reaping the maximum benefits from education. To what extent should our current results be generalized outside of sports? We believe rather broadly; we have no reason to predict that results would differ for teams that rely on intellectual rather than physical performance. According to large meta-analyses, practice effects in intellectual domains tend to be moderate in size (Hausknecht *et al.*, 2007). Even if the overall effect of training is small, we still expect that difficult training helps performance – intellectual as well as physical – more than easy training.

Across various policy areas, such as in education and in the workplace, practitioners need to design training that will best help people succeed. As the nature of work changes, effective policy can help ensure that teams of workers acquire the skills they need and remain resilient in the face of challenge (e.g., Van der Hoek *et al.*, 2018). Previous research in the behavioral sciences, reviewed at the start of this paper, suggested that difficult training is the most useful, but we suspect that this evidence has not yet been fully incorporated into practice (although we have evidence for this gap only in sports). Focusing on the domain of sports, we identify several reasons why this may be the case (e.g., long-term effects, team versus individual performance, the possibility of extreme difficulty backfiring), reject these possibilities following analysis of archival data and show that difficult training helps teams succeed. Interestingly, our survey experiment with university student participants suggests that most people recognize the merits of difficult training for performance – though they may overestimate its dampening effects on motivation – but hesitate to prescribe it to others to the same extent as they would do for themselves. Thus, whereas other research in behavioral science might close by suggesting that people try to gain distance from decisions (e.g., imagine giving advice to someone else rather than deciding for oneself), when it comes to training, we recommend the opposite. Policy-makers should not hesitate to select the same difficult training for others as they would willingly use for themselves.

https://doi.org/10.1017/bpp.2019.22 Published online by Cambridge University Press

Acknowledgements

We are grateful to Gavin Kilduff, Chris To and Severine Toussaert for their helpful comments on this research.

Author note

Data and materials are available at https://osf.io/cr8wx/.

References

Alliger, G. M., C. P. Cerasoli, S. I. Tannenbaum and W. B. Vessey (2015) 'Team resilience', *Organizational Dynamics*, **44**, 176–184.

Alter, A. L., D. M. Oppenheimer, N. Epley and R. N. Eyre (2007) 'Overcoming intuition: metacognitive difficulty activates analytic reasoning', *Journal of Experimental Psychology: General*, **136**, 569.

Bandura, A., (1997) *Self-efficacy: The exercise of control*, Macmillan.

Barch, A. M., D. Lewis (1954) 'The effect of task difficulty and amount of practice on proactive transfer', *Journal of experimental psychology*, **48**, 134.

Bjork, E. L., R. Bjork (2011) 'Making things hard on yourself, but in a good way', *Psychology in the Real World*, 59–68.

Bloom, G., D. Stevens and T. Wickwire (2003) 'Expert Coaches' Perceptions of Team Building', *Journal of Applied Sport Psychology*, **15**, 129–143. https://doi.org/10.1080/10413200305397

Cohen, S. G., D. E. Bailey (1997) 'What makes teams work: Group effectiveness research from the shop floor to the executive suite', *Journal of management*, **23**, 239–290.

Collins, D., Á. MacNamara (2012) The rocky road to the top. Sports medicine 42, 907–914.

Deloitte, (2018) 2018 Global Human Capital Trends.

Drinen, D., (2006) A very simple ranking system » Pro-football-reference.com blog » Blog Archive [www Document]. URL https://www.pro-football-reference.com/blog/index4837.html?p=37.

Duckworth, A. L., C. Peterson, M. D. Matthews and D. R. Kelly (2007) 'Grit: perseverance and passion for long-term goals', *Journal of personality and social psychology*, **92**, 1087.

Fisher, K. M., J. I. Lipson (1986) 'Twenty questions about student errors', *Journal of Research in Science Teaching*, **23**, 783–803.

Garland, H., (1983) 'Influence of ability, assigned goals, and normative information on personal goals and performance: A challenge to the goal attainability assumption', *Journal of Applied Psychology*, **68**, 20.

Gourieroux, C., A. Monfort and A. Trognon (1984) 'Pseudo maximum likelihood methods: Theory', *Econometrica: Journal of the Econometric Society*, 681–700.

Gully, S. M., K. A. Incalcaterra, A. Joshi and J. M. Beaubien (2002) 'A meta-analysis of team-efficacy, potency, and performance: interdependence and level of analysis as moderators of observed relationships', *Journal of applied psychology*, **87**, 819.

Hausknecht, J. P., J. A. Halpert, N. T. Di Paolo and M. O. Moriarty Gerrard (2007) 'Retesting in selection: a meta-analysis of coaching and practice effects for tests of cognitive ability', *Journal of Applied Psychology*, **92**, 373.

Hausman, J. A., B. H. Hall and Z. Griliches (1984) *Econometric models for count data with an application to the patents–R&D relationship*, USA, MA: National Bureau of Economic Research Cambridge.

https://doi.org/10.1017/bpp.2019.22 Published online by Cambridge University Press

Heimbeck, D., M. Frese, S. Sonnentag and N. Keith (2003) 'Integrating errors into the training process: The function of error management instructions and the role of goal orientation', *Personnel Psychology*, **56**, 333–361.

Holding, D. H., (1962) 'Transfer between difficult and easy tasks', *British Journal of Psychology*, **53**, 397–407.

Hsee, C. K., E. U. Weber (1997) 'A fundamental prediction error: Self–others discrepancies in risk preference', *Journal of Experimental Psychology: General*, **126**, 45–53. https://doi.org/10.1037/0096-3445.126.1.45

Katz, N., (2001) 'Sports teams as a model for workplace teams: Lessons and liabilities', *The Academy of Management Executive*, **15**, 56–67.

Keith, N., M. Frese (2008) 'Effectiveness of error management training: a meta-analysis', *Journal of Applied Psychology*, **93**, 59.

Kleingeld, A., H. van Mierlo and L. Arends (2011) 'The effect of goal setting on group performance: A meta-analysis', *Journal of Applied Psychology*, **96**, 1289.

Locke, E. A., G. P. Latham (1990) *A theory of goal setting & task performance*, Prentice-Hall, Inc.

Neri, D. J., (2017) Gridiron Grit: Angela Duckworth on Excelling in the NFL and Beyond. Behavioral Scientist. URL http://behavioralscientist.org/gridiron-grit-angela-duckworth-excelling-nfl-beyond/.

Pitts, J. D., (2016) 'Determinants of Success in the National Football League's Postseason: How Important Is Previous Playoff Experience? *Journal of Sports Economics*, **17**, 86–111.

Seery, M. D., E. A. Holman and R. C. Silver (2010) 'Whatever does not kill us: cumulative lifetime adversity, vulnerability, and resilience', *Journal of Personality and Social Psychology*, **99**, 1025.

Sitkin, S. B., K. E. See, C. C. Miller, M. W. Lawless and A. M. Carton (2011) 'The paradox of stretch goals: Organizations in pursuit of the seemingly impossible', *Academy of Management Review*, **36**, 544–566.

Sitzmann, T., K. Ely, K. G. Brown and K. N. Bauer (2010) 'Self-assessment of knowledge: A cognitive learning or affective measure? *Academy of Management Learning & Education*, **9**, 169–191.

Soderstrom, N. C., R. A. Bjork (2015) 'Learning versus performance: An integrative review', *Perspectives on Psychological Science*, **10**, 176–199.

Van der Hoek, M., S. Groeneveld and B. Kuipers (2018) 'Goal setting in teams: Goal clarity and team performance in the public sector', *Review of public personnel administration*, **38**, 472–493.

Wooldridge, J. M., (1999) 'Distribution-free estimation of some nonlinear panel data models', *Journal of Econometrics*, **90**, 77–97.

Zikmund-Fisher, B. J., B. Sarr, A. Fagerlin and P. A. Ubel (2006) 'A matter of perspective: choosing for others differs from choosing for yourself in making treatment decisions', *Journal of General Internal Medicine*, **21**, 618–622.

https://doi.org/10.1017/bpp.2019.22 Published online by Cambridge University Press